Growth Curve Analysis

by

Seymour Geisser
University of Minnesota

Technical Report No. 324

August 1978

Growth Curve Analysis

by Seymour Geisser
University of Minnesota

1. Repeated Measurements -- the Profile Background

Multiple observations are often made on units or individuals who have been sampled from one or more populations or groups. The observations themselves may be made over differing conditions, tests or periods of time, but all responses are in a comparable metric. The data may be represented by the random vector $X'_{\alpha j} = (X_{1\alpha j}, \ldots, X_{p\alpha j})$, $j = 1, \ldots, N_\alpha$, $\alpha = 1, \ldots, q$ where $X_{i\alpha j} = x_{i\alpha j}$ is the observed response individual in the α^{th} group to the i^{th} stimulus, treatment or to a single treatment at time t_i (we shall subsume all of these possibilities by the term "variables"). Assume that $X_{\alpha j}$ is a multivariate normal random vector with $E(X'_{\alpha j}) = \mu'_\alpha = (\mu_{1\alpha}, \ldots, \mu_{p\alpha})$ and $\mathrm{Cov}(X_{\alpha j}) = \Sigma$. A problem which is often of interest is whether the q groups have parallel profiles; i.e., whether there is a consistency of shape to the mean vectors μ_α. When $\Sigma = \sigma^2 I$, the test of the hypothesis of parallel profiles is equivalent to the test of no interaction of groups by variables which derives from an analysis of variance table. Even for a Σ such that the variance of the difference between any pair of variables is constant, the usual distribution of the analysis of variance F ratio of groups by variables to individuals by variables within groups will still obtain. A special case of this latter condition is the uniform covariance matrix structure $\Sigma = \sigma^2[(1-\rho)I + \rho ee']$, where e is a p-dimensional vector whose components are all unity. The requisite

*The standard notation of the realization of a random variable is too cumbersome to maintain throughout this paper and shall be relaxed in subsequent sections. The meaning, however, will be clear from the context even though rigorous notation is being abused.

sums of squares are given as

$$Q_1 = N \sum_{i=1}^{p} (\bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2, \quad Q_2 = p \sum_{\alpha=1}^{q} N_\alpha (\bar{x}_{\cdot\alpha\cdot} - \bar{x}_{\cdot\cdot\cdot})^2, \quad Q_3 = p \sum_{k=1}^{q} \sum_{j=1}^{N_k} (\bar{x}_{\cdot\alpha j} - \bar{x}_{\cdot\alpha\cdot})^2,$$

$$Q_4 = \sum_{\alpha=1}^{q} \sum_{i=1}^{p} N_\alpha (\bar{x}_{i\alpha\cdot} - \bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\alpha\cdot} + \bar{x}_{\cdot\cdot\cdot})^2, \quad Q_5 = \sum_{\alpha=1}^{q} \sum_{j=1}^{N_\alpha} \sum_{i=1}^{p} (x_{i\alpha j} - \bar{x}_{i\alpha\cdot} - \bar{x}_{\cdot\alpha j} + \bar{x}_{\cdot\alpha\cdot})^2,$$

$$Q_6 = \sum_\alpha \sum_i \sum_j (x_{i\alpha j} - \bar{x}_{\cdot\cdot\cdot})^2, \quad \bar{x}_{i\alpha\cdot} = N_\alpha^{-1} \sum_j x_{i\alpha j}, \quad \bar{x}_{\cdot\alpha j} = p^{-1} \sum_i x_{i\alpha j} \quad \text{and}$$

$$\bar{x}_{\cdot\alpha\cdot} = N_\alpha^{-1} \sum_{j=1}^{N_\alpha} \bar{x}_{\cdot\alpha j}, \quad \bar{x}_{i\cdot\cdot} = q^{-1} \sum_{\alpha=1}^{q} \bar{x}_{i\alpha\cdot}, \quad \bar{x}_{\cdot\cdot\cdot} = N^{-1} \sum_i \sum_\alpha \sum_j x_{i\alpha j}, \quad N = \sum_\alpha N_\alpha$$

The analysis of variance table is then given as follows.

TABLE 1

Analysis of Variance

Source	d.f.	s.s.	F
Variables	p-1	Q_1	$F_1 = (N-q)\dfrac{Q_1}{Q_5}$
Groups	q-1	Q_2	$F_2 = \dfrac{(N-q)Q_2}{(q-1)Q_3}$
Individuals (within Groups)	N-q	Q_3	
Group x Variables	(p-1)(q-1)	Q_4	$F_3 = \dfrac{(N-q)Q_4}{(q-1)Q_5}$
Indiv. x Variables (within Groups)	(p-1)(N-q)	Q_5	
Total	Np-1	Q_6	

Under this restriction on Σ, F_1, F_2 and F_3 are distributed as $F(p-1, (p-1)(N-q))$, $F(q-1,N-q)$ and $F[(p-1)(q-1), (p-1)(N-q)]$ respectively. The test rejects the hypothesis of parallel profiles at level β if $F_3 > F_\beta[(p-1)(q-1), (p-1)(N-q)]$, where $\Pr[F_3 > F_\beta] = \beta$. If the covariance matrix Σ is arbitrary, F_3 was shown to be approximately distributed as F under the hypothesis of no interaction, but with reduced degrees of freedom (Geisser and Greenhouse, 1958, 1959). Here F_3 is approximately $F[(p-1)(q-1)\epsilon, (p-1)(N-q)\epsilon]$ where

$$\epsilon = (p-1)^{-1}[\mathrm{Tr}(\Sigma - p^{-1}ee'\Sigma)]^2/\mathrm{Tr}[\Sigma - p^{-1}ee'\Sigma]^2 \geq (p-1)^{-1} \tag{1.1}$$

and e is a p-dimensional vector of ones. An estimate of ϵ is obtained by replacing Σ with an estimator $\hat{\Sigma} = (N-q)^{-1} \sum_{\alpha=1}^{q} \sum_{j=1}^{N_\alpha} (X_{\alpha j} - \bar{X}_\alpha)(X_{\alpha j} - \bar{X}_\alpha)'$ in the original vectorial representation. The size and power of this test procedure using $\hat{\epsilon}$ has been studied by Collier et. al (1967) and Wilson (1975). Since $\epsilon \geq (p-1)^{-1}$ independently of the form of Σ one may use $F_\alpha(q-1,N-q)$ as a conservative value for F_3. This may be of value in particular when $N - q \leq p$ (where multivariate procedures, to be subsequently discussed, cannot be applied) or when Σ is not the same for each group.

Actually when Σ is assumed arbitrary an exact multivariate test can be made. This is accomplished by eliminating the level of the vector by transforming $Y_{\alpha j} = CX_{\alpha j}$ so that $E(Y_{\alpha j}) = \eta_\alpha = C\mu_\alpha$ where C is any $p-1 \times p$ matrix of rank $p-1$ such that $Ce = 0$ and e is a p-dimensional vector all of whose components are unity. Hence the new p-1 dimensional vectors $\eta_1, \ldots, \eta_q$ are all the same if and only if the parallel profile hypothesis is true. Hence the test of H_0: $\eta_1 = \ldots = \eta_p$ is a one-way multivariate analysis of variance test on the transformed random vectors $Y_{\alpha j}$. The usual test statistic for H_0, for

$$A = \sum_{\alpha=1}^{q} \sum_{j=1}^{N_\alpha} (Y_{\alpha j} - \overline{Y}_{\alpha})(Y_{\alpha j} - \overline{Y}_{\alpha})' \quad \text{and} \quad B = \sum_\alpha N_\alpha(\overline{Y}_\alpha - \overline{Y})(\overline{Y}_\alpha - \overline{Y}), \ \overline{Y} = N^{-1}\sum_\alpha N_\alpha,$$

$\overline{Y}_\alpha = N_\alpha^{-1} \sum_j Y_{\alpha j}$, is

$$|I + A^{-1}B| = U_{p-1,\ q-1,\ N-q} \tag{1.2}$$

where $U_{r,s,t} = \prod_{j=1}^{r} X_j$ for X_j independently distributed as beta

variates with parameters $(t+1-j)/2$ and $s/2$ where a beta density is given as

$$f(x|a,b) = \frac{\Gamma(a + b)}{\Gamma(a)\ \Gamma(b)} x^{a-1}(1-x)^{b-1} \quad .$$

Exact percentage points of the statistic U have been tabled by Schatzoff (1966) and by Lee (1972).

A worked example is presented by Greenhouse and Geisser (1959) employing both techniques and comparing them. Other worked examples appear in Danford, Hughes and McNee (1960) and Cole and Grizzle (1966).

Sometimes a simultaneous region for these parallel profile differentials is of interest. Assume that $\mu_\alpha = \mu_1 + \theta_\alpha e$, where $\theta_1 = 0$; i.e., the parallel profile hypothesis is true. A Bayesian solution to the problem of a simultaneous region for $\theta' = (\theta_2, \ldots, \theta_q)$ is given by Geisser (1965a) and Geisser and Kappenman (1971). They assume the convenient prior density $g(\mu_1, \theta, \Sigma^{-1}) \propto |\Sigma|^{\frac{p+1}{2}}$. This yields the posterior probability statement

$$P\{Q(\theta) \leq F_\beta(q-1,\ N-q)\} = 1 - \beta \tag{1.3}$$

where $F(a,b)$ represents the F distribution with a and b degrees of freedom and

$$Q(\theta) = (q-1)^{-1}(N-q)(e'A^{-1}e)[\theta - (e'A^{-1}e)^{-1}Z'A^{-1}e]' \cdot [\eta^{-1} + Z'A^{-1}Z-(e'A^{-1}e)^{-1}(Z'A^{-1}ee'A^{-1}Z)]^{-1}[\theta-(e'A^{-1}e)^{-1}Z'A^{-1}e] \tag{1.4}$$

where $A = \sum_{\alpha=1}^{q} \sum_{j=1}^{N_\alpha} (X_{\alpha j} - \bar{X}_\alpha)(X_{\alpha j} - \bar{X}_\alpha)'$, $\bar{X}_\alpha = N_\alpha^{-1} \sum_{j=1}^{N_\alpha} X_{\alpha j}$,

$Z = (Z_2, \ldots, Z_q)$, $Z_\alpha = \bar{X}_\alpha - \bar{X}_1$, $\alpha = 2, \ldots, q$; $N = \sum_{\alpha=1}^{q} N_\alpha$

$$\eta = N^{-1} \begin{bmatrix} N_2(N - N_1) & -N_2N_3 \cdots\cdots & \cdots & -N_2N_q \\ & N_3(N - N_3) & -N_3N_4 \cdots & -N_3N_q \\ & & \ddots & \vdots \\ & & & -N_{q-1}N_q \\ & & & N_q(N - N_q) \end{bmatrix}$$

a symmetric matrix. This can easily be extended to natural conjugate prior densities; see Geisser (1965).where a complete analysis was made for $q = 2$. There does not appear to be a confidence region of comparable simplicity, e.g. Halperin (1961).

2. Growth Curve Models.

Originally, the parallel profile problem was subsumed under the general rubric of growth curves by Box (1950). Later Potthof and Roy (1964), Rao (1959, 1965, 1966, 1969) defined the growth curve problem as one in which the components of μ_α were known linear combinations of some subset of unknown parameters. In general then, the column vectors of the $p \times N$ random matrix say, $X = (X_1, \ldots, X_N)$ are assumed to be independently and normally distributed with common covariance matrix Σ and $E(X) = W_{p\times m}\, \tau_{m\times q}\, Z_{q\times N}$ where W is known and of rank $m < p$, Z is known and of rank $q < N$, and τ is unknown. A set of problems involves the estimation and testing of τ and known linear functions of the elements of τ. Although this model was proposed by Potthoff and Roy (1964), their analysis turned out to be inadequate as Rao (1966) demonstrated. This model, however, turned out to be rather fruitful in that it provided a general format for a variety of growth curve situations. In particular, polynomial curves in time as models for growth curves are an important example. This comes about in the following way: Let

$$W = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{m-1} \\ 1 & t_2 & t_2^2 & & t_2^{m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_p & t_p^2 & & t_p^{m-1} \end{pmatrix} \tag{2.1}$$

$$\tau = (\tau_1, \tau_2, \ldots, \tau_q)$$

where $\tau_\alpha' = (\tau_{1\alpha}, \tau_{2\alpha}, \ldots, \tau_{m\alpha}) \qquad \alpha = 1, \ldots, q$.

$$Z = \begin{pmatrix} e_1' & 0_2', \ldots & 0_q' \\ 0_1' & e_2', \; 0_3'. & 0_q' \\ \vdots & \vdots & \vdots \\ 0_1' & \cdots\cdots & e_q' \end{pmatrix} \tag{2.2}$$

e_α is a $N_\alpha \times 1$ vector all of whose components are unity and 0_α is the null vector of size N_α. This yields

$$E(X'_{\alpha j}) = (\tau_{1\alpha} + \sum_{k=2}^{m} \tau_{k\alpha} t_1^{k-1}, \tau_{1\alpha} + \sum_{k=2}^{m} \tau_{k\alpha} t_2^{k-1}, \ldots, \tau_{1\alpha} + \sum_{k=2}^{m} \tau_{k\alpha} t_p^{k-1}) \quad (2.3)$$

e.g., a linear model results from $m = 2$ and

$$E(X'_{\alpha j}) = (\tau_{1\alpha} + \tau_{2\alpha} t_1, \tau_{1\alpha} + \tau_{2\alpha} t_2, \ldots, \tau_{1\alpha} + \tau_{2\alpha} t_p) \quad . \quad (2.4)$$

Further a variety of hypotheses concerning the elements of τ are easily formulated as $C\tau D = 0$ where D is a $q \times d$ matrix of rank $d \le q$ and C is a $c \times m$ matrix of rank $c \le m$. For example, in the previously discussed linear case, one may be only interested in testing H_0: $\tau_{21} = \tau_{22} = \tau_{23} = .. = \tau_{2q}$; i.e., that all the groups "grew" at an equal rate. Hence,

$$0 = C \tau D = (0,1) \begin{pmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1q} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2q} \end{pmatrix} D \quad (2.5)$$

where D is any $q \times q-1$ matrix of rank $q-1$ such that the columns of D sum to zero.

Some formulations involve special structure on Σ. Other formulations depend on hierarchical models.

3. Classical Multivariate Model - Frequentist Analysis

For Σ arbitrary, Rao (1966) demonstrated that the appropriate least squares estimator of τ was

$$\hat{\tau} = (W'A^{-1}W)^{-1}W'A^{-1}XZ'(ZZ')^{-1}$$

where

$$A = X(I-Z'(ZZ')^{-1}Z)X' \quad . \quad (3.1)$$

Khatri (1966) showed that $\hat{\tau}$ was also the maximum likelihood estimator. In the series of papers by Rao (1959, 1965, 1966, 1967) and Khatri (1966), the basic sampling distribution theory was presented. A $1-\beta$ confidence region on τ is found from

$$\Pr[Q(\tau) \geq U_\beta] = 1 - \beta$$

where

$$Q(\tau) = |I + W'A^{-1}W(\hat{\tau} - \tau)G(\hat{\tau} - \tau)'|^{-1} \tag{3.2}$$

and

$$G^{-1} = (ZZ')^{-1} + (ZZ')^{-1}ZX'[A^{-1} - A^{-1}W(W'A^{-1}W)^{-1}W'A^{-1}]XZ'(ZZ')^{-1} \tag{3.3}$$

and U_β is the βth percentage point such that

$$\Pr[U_{m,q,N-q+m-p} \geq U_\beta] = 1 - \beta .$$

The null hypothesis that $\tau = \tau_o$ is rejected at level β if $Q(\tau_o) \leq U_\beta$.

Confidence regions for a variety of linear combinations of the elements of τ can be obtained by noting that

$$Q(C\tau D) = |I + [C(W'A^{-1}W)^{-1}C']^{-1}(C\hat{\tau}D-C\tau D)(D'G^{-1}D)^{-1}(C\hat{\tau}D-C\tau D)'|^{-1} \tag{3.4}$$

is distributed as $U_{c,d,N-q-p+m}$. Many useful null hypotheses, as indicated before, can be expressed as $C\tau D = 0$, for appropriate C and D .

In the simplest case where we are dealing with one group i.e., $q = 1$ and $Z = (1, \ldots, 1)$.

$$R(\tau) = \frac{N(\hat{\tau}-\tau)'W'A^{-1}W(\hat{\tau}-\tau)}{1+N\,T_2'(U'AU)^{-1}T_2} \sim m(N-p)^{-1}F(m,N-p) \tag{3.5}$$

and U is any $p \times p-m$ matrix of rank $p-m$ such that $U'W = 0$ and $T_2=U'XZ'(ZZ')^{-1}$.

Hence a $1-\beta$ hyperellipsoidal confidence region for the m dimensional vector τ is obtained from $F_\beta(m,N-p)$ the β^{th} percentage point so that all τ satisfying

$$\Pr[R(\tau) \leq F_\beta(m,N-p)] = 1-\beta$$

are included in the region.

Before this type of analysis was introduced it was well known that the statistic

$$T_1 = B\,X\,Z'(ZZ')^{-1}, \tag{3.6}$$

for $B=(W'W)^{-1}W'$, was an unbiased estimator of τ and that a confidence region for τ could be obtained from

$$Q(\tau) = |I + (BAB')^{-1}(T_1 - \tau)ZZ'(T_1 - \tau)'|^{-1} \tag{3.7}$$

which is distributed as $U_{m,q,N-q}$. The form for $C\tau D$, analogous to (3.4), is

$$Q(C\tau D) = |I + (CBAB'C')^{-1}(CTD-C\tau D)(D'(ZZ')^{-1}D)^{-1}(CT_1D-C\tau D)'|^{-1} \tag{3.8}$$

which is distributed as $U_{c,d,N-q}$.

From 3.7 $q = 1$ and $Z = (1, \ldots, 1)$ we obtain

$$R(\tau) = N(T_1 - \tau)'(BAB')^{-1}(T_1 - \tau) \sim m(N-m)^{-1}F(m,N-m)\ . \tag{3.9}$$

Further one can write

$$\hat{\tau} = T_1 - BAU(U'AU)^{-1}T_2$$

with T_2 and U as previously defined. This displays the fact that $\hat{\tau}$ is a covariance adjusted estimator. Since both $E(T_1) = E(\hat{\tau}) = \tau$, then comparisons of their covariance matrices would be instructive as to which would be a more desireable estimator. It turns out that T_1 is preferable when

$$B\,\Sigma\,U = 0 \tag{3.10}$$

and possibly when this matrix is close to the null matrix, otherwise $\hat{\tau}$ is apparently preferable. For $\Sigma = \sigma^2 I$ (3.10) certainly holds. More generally it will hold for

$$\Sigma = W\Gamma W' + U\theta U' + \sigma^2 I\ . \tag{3.11}$$

In fact, Rao (1967, 1968) shows that if and only if (3.11) holds then T_1 is the least squares estimator of τ.

A likelihood ratio test for

$$H_0: \ \Sigma = W\Gamma W' + U\theta U' + \sigma^2 I$$

vs.

$$H_1: \ \Sigma \neq W\Gamma W' + U\theta U' + \sigma^2 I$$

is easily obtained. The test statistic

$$\lambda = |W'A^{-1}WBAB'|$$

for testing H_0 vs. H_1 is distributed as $U_{m,p-m,N-p-1+m}$ under H_0, c.f. Lee and Geisser (1972).

Other models for Σ that have been studied are the factor analytic model of Rao (1967)

$$\Sigma = C\Gamma C' + \sigma^2 I \tag{3.12}$$

and the serial correlation model

$$\Sigma = \{\sigma_{ij}\} = \{\sigma^2\rho^{|i-j|}\} \quad i,j = 1, \ldots p . \tag{3.13}$$

but optimal results for estimation are difficult to achieve.

In some instances a confidence region either on a particular point of the growth curve or on the entire growth curve itself is of interest. Suppose W is of the form (2.1) i.e. the growth curve is polynomial and Σ is arbitrary. Then let $C = a' = (1, t, t^2, \ldots, t^{m-1})$ and $D = I$ so that for a given value of t

$$C\tau D = (1, t, \ldots, t^{m-1})(\tau_1, \ldots, \tau_q) = (a'\tau_1, \ldots, a'\tau_q) .$$

One then applies (3.4) which reduces to

$$Q(a'\tau) = |I + [a'(W'A^{-1}W)^{-1}a]^{-1} (a'\hat{\tau}-a'\tau)G(a'\hat{\tau}-a'\tau)'|^{-1}$$

distributed as $U_{1,q,N-q-p+m}$. But since

$$U_{1,q,N-q-p+m} = (1 + q(N-q-p+m)^{-1})F(q,N-q-p+m)$$

then

$$\frac{(a'\hat{\tau} - a'\tau)G(a'\hat{\tau} - a'\tau)'}{a'(W'A^{-1}W)^{-1}a} \sim \frac{q}{N-q-p+m} F(q,N-q-p+m) \qquad (3.14)$$

which provides a joint confidence region on the q polynomials at a given value of t. If $q = 1$ so that only one group and one polynomial is involved then $\tau = \tau_1$ and

$$\frac{N(a'\hat{\tau} - a'\tau)'(a'\hat{\tau} - a\tau)}{(a'(W'A^{-1}W)^{-1}a)(1 + NT_2'(U'AU)^{-1}T_2)} \sim (N-1-p+m)^{-1}F(1,N-1-p+m). \qquad (3.15)$$

Note that $N^{-1} + T_2'(U'AU)^{-1}T_2 = G^{-1}$ can be given without computing U by applying (3.3).

A simultaneous confidence region for the entire growth curve i.e. for all t, is obtained by noting that

$$\Pr\left[\frac{N(N-1-p+m)(a'\hat{\tau} - a'\tau)^2}{m(1+NT_2'(U'AU)^{-1}T_2)a'(W'A^{-1}W)^{-1}a} \le F_\beta(m,N-p)\right] \ge 1-\beta \ , \qquad (3.16)$$

When Σ is of the form (3.13) then similar results are obtainable so that analogous to (3.15) we obtain

$$\frac{N(\hat{a}T_1 - a'\tau)'(a'T_1 - a\tau)}{a'BAB'a} \sim (N-1)^{-1}F(1,N-1) \qquad (3.17)$$

for a single value of t. For a simultaneous region for the entire growth curve we can use the fact that

$$\Pr\left[\frac{N(N-1)(a'T_1 - a'\tau)^2}{m \quad a'BAB'a} \le F_\beta(m,N-1)\right] \ge 1 - \beta \ . \qquad (3.18)$$

Suppose a tolerance region is required on k future p-dimensional independent multivariate normal vectors with common covariance matrix Σ. Denote this set of variables by V_{pxK} and assume $E(V_{pxK}) = W\tau F$ where F is a $q \times K$ known design matrix. It can easily be shown that for $H = (Z,F)$

$$U = |I + (I-F'(HH')^{-1}F)(V-WT_1F)'W\,(W'AW)^{-1}W'(V-WT_1F)|^{-1} \qquad (3.19)$$

is distributed as $U_{m,K,N-q}$ irrespective of the form of Σ, Geisser (1970)

For $K = F \doteq q = 1$ and $Z = (1, \ldots, 1)$ i.e. a single vector observation to be predicted from a single group of observations,

$$y_1 = \frac{N}{N+1} (V - WT_1)'W (W'AW)^{-1}W'(V - WT_1) \tag{3.20}$$

is distributed as $m(N-m)^{-1}F(m,N-m)$.

It is clear that using (3.20) or (3.19) for a tolerance region would be unsatisfactory since $W (W'AW)^{-1}W'$ is singular. However, we also note that, independently of U_1,

$$U_2 = |I + (V-WT_1F)'U(U'XX'U)^{-1}U'(V-WT_1F)|^{-1} \tag{3.21}$$

is distributed as $U_{p-m,K,N}$. For the case corresponding to (3.20),

$$y_2 = (V-WT_1F)'U(U'XX'U)^{-1}U'(V-WT_1F) \tag{3.22}$$

is distributed as $(p-m)(N+m+1-p)^{-1}F(p-m,N+m+1-p)$. Hence for general V, a tolerance region can be obtained from the distribution of $U_1 + U_2$ or for the special case

$$y = y_1 + y_2 = (V-WT_1)'[N(N+1)^{-1}W (W'AW)^{-1}W'+U(U'XX'U)^{-1}U'](V-WT_1) \tag{3.23}$$

is distributed as a linear sum of two independent F variates. Now the matrix of the quadratic form is positive definite with probability one so that a $1 - \beta$ hyperellipsoidal tolerance region emerges from the observed X which includes all V such that $y \leq y_\beta$ where y_β is the β^{th} percentage point of the linear sum of independent F variables.

Now if Σ is of the specialized form (3.13) where T_1 is the optimal estimator of τ , the above tolerance region would undoubtedly enjoy its "best" coverage properties. However, if Σ is not of this form the above tolerance region need no longer exhibit as good coverage properties as some other one.

For the arbitrary Σ , it can be shown that

$$U_1 = |I + G_1(V-W\hat{\tau}F)'A^{-1}W(W'A^{-1}W)^{-1}W'A^{-1}(V-W\hat{\tau}F)|^{-1} \sim U_{m,K,N-q+m-p} \quad (3.24)$$

where

$$G_1^{-1} = (I-F'(HH')^{-1}F)^{-1} + (V-XZ'(ZZ')^{-1}F)'U(U'AU)^{-1}U'(V-XZ'(ZZ')^{-1}F)$$

and $E = (Z,F)$. For $K = F = q = 1$ and $Z = (1, \ldots, 1)$ and $\bar{X}$ the sample mean vector,

$$y_1 = \frac{N(N+1)^{-1}(V-W\hat{\tau})'A^{-1}W(W'A^{-1}W)^{-1}W'A^{-1}(V-W\hat{\tau}F)}{1 + N(N+1)^{-1}(V-\bar{X})'U(U'AU)^{-1}U'(V-\bar{X})} \sim m(N-p)F(m,N-p) \quad (3.25)$$

Now independent of U_1 ,

$$U_2 = |I + (V - W\hat{\tau}F)'U(U'XX'U)^{-1}U'(V - W\hat{\tau}F)|^{-1} \quad (3.26)$$

is distributed as $U_{p-m,K,N}$. For the case corresponding to (3.25)

$$y_2 = (V-W\hat{\tau} F)'U(U'XX'U)^{-1}U'(V-W\hat{\tau}F) \sim \frac{(p-m)}{N+m+1-p} F(p-m,N+m+1-p) \; . \quad (3.27)$$

Therefore, as before, a tolerance region for V can, in theory, be obtained from the distribution of $U_1 + U_2$ or for the special case $y = y_1 + y_2$. However, the tolerance region apparently will include V in disconnected regions due to the form of y_1 . What properties this type of coverage will possess is not clear at present.

Another problem of great interest brought to the fore by Lee and Geisser (1972, 1975) is the problem of conditional predictive or tolerance regions.

Suppose V can be partitioned into an observed portion $V^{(1)}$ which is $p_1 \times K$ and an unobserved portion $V^{(2)}$ which is $p_2 \times K$ such that

$$V = \begin{pmatrix} V^{(1)} \\ V^{(2)} \end{pmatrix}$$

and $p_1 + p_2 = p$. A tolerance region for $V^{(2)}$ is required for the observed $V^{(1)}$ and X. It would appear that the results previously obtained from (3.19) and (3.20) can be used, as before, except that now values are also inserted for $V^{(1)}$ and the region consists of all $V^{(2)}$ satisfying $U_1 + U_2 \geq$ const. We shall illustrate this with the case that usually occurs in such problems, $K = F = q = 1$, $Z = (1, \ldots, 1)$. Let

$$N(N+1)^{-1}B'(BAB')^{-1}B+U(U'XX'U)^{-1}U' = P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}^{-1}$$

and

$$W = \begin{pmatrix} W^{(1)} \\ W^{(2)} \end{pmatrix}$$

where P_{ij} and Q_{ij} are $p_i \times p_j$ and $W^{(i)}$ is $p_i \times m$, for $i,j = 1,2$. Then from (3.23) the $1-\beta$ tolerance region consists of all $V^{(2)}$ satisfying

$$y = [V^{(2)}-W^{(2)}T_1-Q_{21}Q_{11}^{-1}(V^{(1)}-W^{(1)}T_1)]'P_{22}[V^{(2)}-W^{(2)}T_1-Q_{21}Q_{11}^{-1}(V^{(1)}-W^{(1)}T_1)]$$
$$+ (V^{(1)}-W^{(1)}T_1)'Q_{11}^{-1}(V^{(1)}-X^{(1)}T_1) \leq y_\beta \qquad (3.28)$$

As is sometimes the case for the confidence region setup, a particular $1 - \beta$ region can exhibit peculiar properties, here we note that if

$$y_\beta \leq (V^{(1)} - W^{(1)}T_1)'Q_{11}^{-1}(V^{(1)} - W^{(1)}T_1)$$

the region for $V^{(2)}$ is empty. Clearly to avoid such a difficulty it would be much more sensible for the tolerance interval to be obtained from a statistic conditional on the observed $V^{(1)}$. However exact conditional tolerance regions do not seem to be feasible either for this model or for the model which results in $\hat{\tau}$ as an optimal estimator for τ. We shall take this up again in the latter part of the next section when discussing a Bayesian approach to conditional prediction of $V^{(2)}$ given $V^{(1)}$.

4. A Bayesian Approach--Estimation

The Bayesian approach presented in this section was given in a series of papers; Geisser (1970), Lee and Geisser (1972, 1975).

Consider the previous model with arbitrary covariance matrix Σ and a convenient prior density, Geisser and Cornfield (1963), Geisser (1965)

$$g(\tau,\Sigma^{-1})\propto|\Sigma|^{(p+1)/2} \quad . \tag{4.1}$$

From (4.1) we obtain the posterior marginal density of τ

$$p(\tau) \propto |(W'A^{-1}W)^{-1} + (\tau-\hat{\tau})G(\tau-\hat{\tau})'|^{-N/2} \quad , \tag{4.2}$$

Geisser (1970). For a posterior region for τ,

$$Q(\tau) = |I + W'A^{-1}W(\tau-\hat{\tau})G(\tau-\hat{\tau})'|^{-1} \sim U_{m,q,N-q} \tag{4.3}$$

and for $C\tau D$

$$Q(C\tau D) = |I + [C(W'A^{-1}W)^{-1}C']^{-1}(C\hat{\tau}D-C\tau D)(D'G^{-1}D)^{-1}(C\hat{\tau}D-C\tau D')'|^{-1} \tag{4.4}$$

is distributed as $U_{c,d,N-q}$ which differs from its frequency distribution $U_{c,d,N-q-p+m}$ given by (3.4). Hence Bayesian regions at a given point t of the polynomial $a'\tau$ and simultaneously for all t can easily be given in a fashion analogous to (3.14-3.18).

For the case analogous to (3.5) we obtain

$$R(\tau) = \frac{N(\hat{\tau}-\tau)'W'A^{-1}W(\hat{\tau}-\tau)}{1+NT_2'(U'AU)^{-1}T_2} \sim \frac{m}{N-m} F(m,N-m) \tag{4.5}$$

Since $(N-m)^{-1}F_\beta(m,N-m) \leq (N-p)^{-1}F_\beta(m,N-p)$ (4.6)

the region given by (4.5) will always be included in (3.5).

The posterior distribution of τ is the general determinantal density, c.f. Geisser (1966), denoted by $D(\cdot|\hat{\tau}, G, (W'A^{-1}W)^{-1}, N)$, where say Y, the random matrix, is distributed as $D(\cdot|\Delta,\Lambda,\Sigma,N)$ if

$$d(Y) = \frac{C_{m,\nu}\pi^{\frac{-q-m}{2}} |\Sigma|^{\nu/2}|\Lambda|^{m/2}}{C_{m,n}|\Sigma+(Y-\Delta)\Lambda(Y-\Delta)'|^{N/2}} \tag{4.7}$$

where Σ is $m \times m$ and p.d. Λ is $q \times q$ and p.d., Y and Δ are $m \times q$ and $\nu = N - q \geq m \geq 1$ and

$$C^{-1}_{m,\nu} = \pi^{m(m-1)/4} \prod_{i=1}^{m} \left(\frac{\nu+1-i}{2}\right) . \tag{4.8}$$

Hence $E(\tau) = \hat{\tau}$, when it exists and $\hat{\tau}$ is also the mode. The posterior expectation of Σ is given as

$$E(\Sigma) = (N-p-1)^{-1}\{(X-W\hat{\tau}Z)(X-W\hat{\tau}Z)' + (N-m-q-1)^{-1}W(W'A^{-1}W)^{-1}W'[\mathrm{tr}G^{-1}ZZ']\} \tag{4.9}$$

Further it can be shown that $E(\Sigma) - \hat{\Sigma}$ is always negative definite where the m.l.e.

$$\hat{\Sigma} = N^{-1}(X - W\hat{\tau}Z)(X-W\hat{\tau}Z)' . \tag{4.10}$$

From this Bayesian viewpoint i.e. Σ arbitrary, it is also possible to obtain T_1 as the posterior expectation. Let the model be given as

$$E(X) = (W,U)\binom{\tau}{\eta}Z = W\tau Z + U\eta Z \tag{4.11}$$

with prior density for Σ^{-1}, τ, η

$$g(\Sigma^{-1},\tau,\eta) \propto |\Sigma|^{(p+1)/2} \quad . \tag{4.12}$$

Then the marginal density of τ is

$$p(\tau) \propto |BAB' + (\tau-T_1)ZZ'(\tau-T_1)'|^{-(N-p+m)/2} \tag{4.13}$$

so that the center of this distribution is T_1 . Hence for q = 1, Z=(1, ...,1)

$$N(\tau-T_1)'(BAB')^{-1}(\tau-T_1) \sim \frac{m}{N-p} F(m,N-p) \ . \tag{4.14}$$

This is to be contrasted with the previous case (4.5) which is essentially conditional on $\eta = 0$.

If we consider the special covariance model $\Sigma = W\Gamma W' + U\theta U' + \sigma^2 I$, we note that $\sigma^2[W(W'W)^{-1}W' + U(U'U)^{-1}U'] = \sigma^2 I$, so that from the frequency point of view, $\Sigma = W[\Gamma+\sigma^2(W'W)^{-1}]W' + U[\theta + \sigma^2(U'U)^{-1}]U'$ where the brackets may be relabeled Γ and θ respectively, but both now p.d. -- no essential difference ensues. From the Bayesian point of view the newly labeled Γ and θ are now dependent on W which some Bayesians find undesirable but we shall regard it as a convenience to do so since the ensuing results are greatly simplified by this device.

Choosing the simple structure model

$$\Sigma = W\Gamma W' + U\theta U' \tag{4.15}$$

and the convenient prior density

$$g(\Gamma^{-1},\theta^{-1},\tau) \propto |\Gamma|^{(m+1)/2} g(\theta^{-1}) \tag{4.16}$$

where $g(\theta^{-1})$ is arbitrary, the posterior density of τ is obtained as

$$p(\tau) \propto |BAB' + (\tau-T_1)ZZ'(\tau-T_1)'|^{-N/2} \ , \tag{4.17}$$

yielding T_1 as the center of the posterior distribution of τ . Hence a posteriori

$$|I + (BAB')^{-1}(\tau-T_1)ZZ'(\tau-T_1)'| \sim U_{m,q,N-q} \qquad (4.18)$$

precisely as its sampling counterpart, (3.7). Regions for $C\tau D$ are readily obtainable in a form equivalent to (3.8).

Further, if

$$g(\theta^{-1}) \propto |\theta|^{(p-m+1)/2}$$

then

$$E(\Gamma) = (N-q-m-1)^{-1}BAB' \qquad (4.19)$$
$$E(\theta) = (N-p+m-1)^{-1}(U'U)^{-1}U'XX'U(U'U)^{-1} \quad .$$

5. Bayesian Prediction -- Simple Structure

If we wish to predict V when the simple structure (4.15) for Σ obtains, the predictive density of V is given as

$$f(V) \propto |U'(XX' + (V-WT_1F)(V-WT_1F)')U|^{-(N+K)/2} \qquad (5.1)$$
$$\cdot\ |I + (I-F'(HH')^{-1}F)(V-WT_1F)'B'(BAB')^{-1}B(V-WT_1F)|^{-(N+K-q)/2}$$

A maximal mode of the density of V is clearly WT_1F which is also its predictive expectation. It can be easily shown that a posteriori

$$U_1 = |I + (I-F'(HH')^{-1}F)(V-WT_1F)'W(W'AW)^{-1}W'(V-WT_1F)|^{-1} \qquad (5.2)$$

is distributed $U_{m,K,N-q}$ exactly as its frequency distribution given in (3.19). Further a posteriori

$$U_2 = |I + (V-WT_1F)'U(U'XX'U)^{-1}U'(V-WT_1F)|^{-1} \qquad (5.3)$$

is distributed as $U_{p-m,K,N}$ independently of U_1 just as its sampling distribution. Hence the sampling theory given in the previous section e.g. the region generated by the statistic (3.23) has a Bayesian interpretation.

Sometimes for the sake of comparisons, linear combinations of the individual p-dimensional future vectors of $V = (V_1, \ldots, V_k)$ are of interest. Previously the comparisons were made parametrically via regions on $C\tau D$. It is often the case that a more informative comparison is vested in predicting a function of the vectors of V, (Geisser (1971)). For example, if one had 2 groups i.e. $q = 2$ then one could obtain the predictive distribution of $V_1 - V_2$ as a comparison rather than, say, the posterior distribution of $\tau_1 - \tau_2$. At any rate if ℓ is a $k \times 1$ arbitrary real non-zero vector then $v = V\ell = \sum_{i=1}^{k} \ell_i V_i$ is a linear combination of $V_1, V_2, \ldots, V_k$ where $\ell' = (\ell_1, \ldots, \ell_k)$.

The predictive density of $V\ell = v$ is obtained as

$$f(v) \propto |U'(XX' + (\ell\ell')^{-1}vv')U|^{-(N+1)/2}$$

$$\cdot \, |\ell'(I-F'(HH')^{-1}F)^{-1}\ell+(v-WT_1F\ell)'W(W'AW)^{-1}W'(v-WT_1F\ell)|^{-(N+1-q)/2} \tag{5.4}$$

Regions similar to (3.23) can now be generated for v. Marginal densities for a subset of the p components of v can also easily be obtained.

For conditional prediction of $V^{(2)}$ when $V^{(1)}$ is observed we note that $f(V^{(2)}|V^{(1)}) \propto f(V)$. However, this still does not resolve the problem of a convenient region on $V^{(2)}$ given $V^{(1)}$. It was shown by Lee and Geisser (1972) that for $K = 1$ the predictive distribution of $V^{(2)}$ given $V^{(1)}$ can be reasonably well approximated as

$$F(V^{(2)}|V^{(1)}) \approx St.(\cdot;\ \mu_{s2\cdot 1}, b(2N+2-q-p_q)J_{22}^{-1}, 2N+2-q) \tag{5.5}$$

where $Y \sim St.(\cdot,\mu,\Sigma,N)$ is a multivariate student distribution with density

$$f(Y) \propto [1 + (Y-\mu)'(\nu\Sigma)^{-1}(Y-\mu)]^{-N/2} \quad , \tag{5.6}$$

a special case of the general determinantal density (4.?); where

$$t \doteq \frac{N+1-p+m}{N-1-p+m}$$

$$\mu_{s2\cdot 1} = W^{(2)}T_1F - J_{22}^{-1}J_{21}(V^{(1)} - W^{(1)}T_1F) \tag{5.7}$$

$$b = 1 + t(N+1-q-m)(N+1-p+m)^{-1} + (V^{(1)}-W^{(1)}T_1F)'J_{11\cdot 2}(V^{(1)}-W^{(1)}T_1F)$$

$$\frac{t(N+1-q-m)}{(N+1-p+m)}(I + F'(ZZ')^{-1}F)^{-1}W(W'AW)^{-1}W' + U(U'XX'U)^{-1}U' = J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} .$$

Hence an approximate predictive region for $V^{(2)}$ given $V^{(1)}$ is obtained from

$$\frac{2(N+2-q-p_2)}{p_2}\,(V^{(2)}-\mu_{s2\cdot 1})'[b^{-1}J_{22}](V^{(2)}-\mu_{s2\cdot 1}) \approx F(p_2, 2N+2-q-p_2) . \tag{5.8}$$

Note for $q = 1 = F$, $Z = (1, \ldots, 1)$ some simplification occurs in that $I + F'(ZZ')^{-1}F = (N + 1)$. At any rate a sensible region results. Numerical procedures are given for $K = 1$ by Lee and Geisser (1972) for obtaining the predictive mode of $V^{(2)}$ given $V^{(1)}$ and an exact solution for the particularly interesting case $p_2 = 1$.

6. Bayesian Prediction -- Arbitrary Covariance Case

For the arbitrary covariance case the predictive distribution of V was obtained by Geisser (1970) as

$$f(V) \propto |U'(XX' + (V-W\hat{\tau}F)(V-W\hat{\tau}F)')U|^{-(N+K-m)/2}$$

$$\cdot |G_1|^{m/2}|I + G_1(V-W\hat{\tau}F)'A^{-1}W(W'A^{-1}W)^{-1}W'A^{-1}(V-W\hat{\tau}F)|^{-(N+K-q)/2} \qquad (6.1)$$

While $E(V) = X\hat{\tau}F$ is easily derived, it is clear from the form of (6.1) that $X\hat{\tau}F$ is not the mode of V . Numerical procedures for calculating the predictive mode when $K = 1$ are given by Lee and Geisser (1972).

For $Q = BXZ'(ZZ')^{-1}F+BAU(U'AU)^{-1}U'(V-XZ'(ZZ')^{-1}F)$,

$$U_1 = |I + (W'A^{-1}W)(BV-Q)G_1(BV-Q)'| \qquad (6.2)$$

is $U_{m,K,N-q}$ and is independent of

$$U_2 = |I + V'U(U'XX'U)^{-1}U'V)| \sim U_{p-m,K,N-m} \;. \qquad (6.3)$$

Hence the predictive distributions of U_1 and U_2 differ from their sampling distributions given by (3.24) and (3.26). Again $U_1 + U_2$ is the sum of two independent U variates so that conceptually a predictive region can be obtained. For $r = K = F = 1$, y_1, as defined by (3.25), is distributed as $m(N-m)^{-1}F(m,N-m)$ and is independent of y_2 as defined by (3.27), which is distributed as $(p-m)(N+1-p)^{-1}F(p-m,N+1-p)$. As noted before these predictive distributions differ slightly from their sampling distributional counterparts. Similarly a disconnected region for V can be obtained from $y = y_1 + y_2$, which may not only be difficult to compute but also not very appealing. For very large samples multivariate normal theory can be applied as a convenient approximation. For $v = V\ell$, similar remarks hold.

For predicting $V^{(2)}$ given $V^{(1)}$ we again note that $f(V^{(2)}|V^{(1)}) \propto f(V)$ but the only situations where this is easily utilized is $K = 1$ and p_2 is low dimensional, 1 or 2 perhaps; thus permitting the density for $V^{(2)}$ given $V^{(1)}$ to be easily plotted. Otherwise there doesn't seem to be any optimal way to utilize the results in an exact manner other than rather crude normal approximations. For $K = 1$, numerical procedures are given by Lee and Geisser (1972) for calculating the conditional predictive mode of $V^{(2)}$.

Lee and Geisser (1975) also examined and compared a wide variety of conditional prediction procedures on two sets of growth curve data. The conclusions reached for those data sets were that the form of Σ was important in reducing predictive error (in particular serial correlation models for Σ) and that growth curves are often highly individual so that past data on completed growth curves may be relatively unimportant for predicting an individual's future growth when compared to his own past data. This means that other models emphasizing individual curves may be useful in many situations.

The predictive distribution of V can also be used for ascertaining which of possibly q growth curve models is most appropriate for V. Synthesizing the Bayesian notions developed for classification, Geisser (1964), and growth curves, Geisser (1970), Lee (1977) provides solutions for this problem. Details are provided for a variety of cases involving various degrees of knowledge about τ_α and Σ_α the parameters of the α^{th} group. The results are then couched as the predictive probability that V belongs to a particular family of growth curves based on prior probabilities of this event. Calculations are made both for the arbitrary case and when Σ_α is of simple structure. This work has been further extended by Nagel and DeWaal (1978).

7. Individual Growth Curves

Suppose now each vector X_α has its own growth curve so that the model is

$$E(X_\alpha) = W\tau_\alpha \qquad \alpha = 1, \ldots, q \quad .$$

We now make certain simplifying assumptions, Fearn (1975), namely that the model is a two stage hierarchial one so that

$$\begin{aligned} X_\alpha | \tau_\alpha, \sigma^2 &\sim N(W\tau_\alpha, \sigma^2 I) \\ \tau_\alpha | \bar{\tau}, \sigma^2, C &\sim N(\bar{\tau}, \Gamma) \quad . \end{aligned} \tag{7.1}$$

Hence the marginal distribution of X_α is easily obtained, to wit:

$$X_\alpha \sim N(W\bar{\tau}, W\Gamma W' + \sigma^2 I) \tag{7.2}$$

This model was considered by Rao (1965, 1967, 1972, 1975) Lee and Geisser (1975) and Fearn (1975). We shall present the analysis of Fearn which is Bayesian and inspired by Lindley and Smith (1972) and Smith (1973).

The posterior distribution of τ_α given X, σ^2 Γ is normal with mean and covariance matrix

$$\begin{aligned} \tau^*_\alpha &= E(\tau_\alpha) = \Omega\hat{\tau}_\alpha + (I-\Omega)q^{-1}\sum_{k=1}^{q}\hat{\tau}_k \\ D^*_\alpha &= \mathrm{Cov}(\tau_\alpha) = \{\Omega + q^{-1}(I-\Omega)\}\sigma^2(W'W)^{-1} \end{aligned} \tag{7.3}$$

respectively where

$$\begin{aligned} \Omega &= (\sigma^{-2}W'W + \Gamma^{-1})^{-1}\sigma^{-2}(W'W) \\ \hat{\tau}_\alpha &= (W'W)^{-1}W'X_\alpha \quad . \end{aligned} \tag{7.4}$$

Further for a uniform prior distribution on $\bar{\tau}$ given σ^2 and Γ the posterior distribution of $\bar{\tau}$ is normal with mean and covariance matrix given by

$$\bar{\tau}^* = E(\bar{\tau}) = q^{-1} \sum_{\alpha=1}^{q} \tau^*_\alpha$$

$$D^* = \mathrm{Cov}(\bar{\tau}^*) = q^{-1}\{\sigma^2(W'W)^{-1}+\Gamma\} \tag{7.5}$$

It is now further assumed that the prior density of σ is $g(\sigma) \propto \sigma^{-1}$ and that Γ^{-1} has a Wishart distribution with ρ degrees of freedom: and matrix R

$$g(\Gamma^{-1}|\rho,R) \propto |\Gamma|^{\frac{-(\rho-m-1)}{2}} \exp\{-\tfrac{1}{2}\mathrm{tr}\Gamma^{-1}R\} . \tag{7.6}$$

Fearn suggests $\rho = m$ as appropriate when knowledge about the precision of Γ is vague. But a value for R (perhaps diagonal in certain cases) gleaned from some prior knowledge is required. Even so the integrations necessary to obtain the appropriate marginal densities are rather difficult. Hence, the following estimates are used as approximations for σ^2 and Γ respectively.

$$\hat{\sigma}^2 = (q(p-m) + 2)^{-1} \sum_{\alpha=1}^{q} (X_\alpha - W\hat{\tau}_\alpha)'(X_\alpha - W\tau_\alpha)$$

$$\hat{\Gamma} = [\sum_{\alpha=1}^{q} (\hat{\tau}_\alpha - \hat{\bar{\tau}})(\hat{\tau}_\alpha - \hat{\bar{\tau}})' + R]/(q-2) . \tag{7.7}$$

These are then inserted in (7.3) and (7.5) to obtain approximate regions for τ_α and $\bar{\tau}$ based on normal approximations. The nominal $1-\beta$ probability for such a region will undoubtedly overestimate the actual value.

At any rate an approximate $1-\beta$ probability region on an individual τ_α is given by

$$(\tau_\alpha - \tau^*)' D^{*-1}_\alpha (\tau_\alpha - \tau^*) \leq \chi^2_\beta(m) \tag{7.8}$$

where $\chi^2_\beta(r)$ represents the β-th percentage point of a chi-squared random variable with r degrees of freedom. Similarly

$$(\bar{\tau}-\bar{\tau}^*)' D^{*-1}(\bar{\tau}-\bar{\tau}^*) \leq \chi^2_\beta(m) \tag{7.9}$$

yields an approximate $1-\beta$ probability region for $\bar{\tau}$. If one wants to estimate a polynomial growth curve e.g. $a'\tau_\alpha$ where $a' = (1,t,t^2,\dots, t^{m-1})$

at a particular value t , then an approximate $1-\beta$ probability region is found from

$$\frac{(a'\tau_\alpha - a'\tau^*)^2}{a'D^*_\alpha a} \leq \chi^2_\beta(1) \ . \tag{7.10}$$

For a simultaneous region on the entire individual growth curve whose probability is approximately at least as large as $1-\beta$ we obtain

$$\frac{(a'\tau_\alpha - a'\tau^*)^2}{ma'D^*_\alpha a} \leq \chi^2_\beta(m) \ . \tag{7.11}$$

Similar results are obtained for the group mean growth curve i.e. at a single point t a $1-\beta$ probability region is obtained from

$$\frac{(a'\bar{\tau} - a'\bar{\tau}^*)^2}{a'D^*a} \leq \chi^2_\beta(1) \tag{7.12}$$

and for all t a region of probability approximately at least as great as $1-\beta$ is

$$\frac{(a'\bar{\tau} - a'\bar{\tau}^*)^2}{ma'D^*a} \leq \chi^2_\beta(m) \ . \tag{7.13}$$

In all cases (7.9-7.13), it is expected that the probability for the given regions are somewhat less than the stipulated $1-\beta$, due to the approximations involved, unless p and q are quite large relative to m .

For some work on tolerance regions based on a frequency approach, the reader is referred to Bowden and Steinhorst (1973).

Consider now predicting a new vector V which is distributed as

$$V | \tau_{q+1}, \sigma^2 \sim N(W\tau_{q+1}, \sigma^2 I)$$

and

$$\tau_{q+1} | \bar{\tau}, \sigma^2, \Gamma \sim N(\bar{\tau}, W\Gamma W' + \sigma^2 I) \ .$$

Now as before the posterior distribution of $\bar{\tau}$, when the prior distribution is uniform given σ^2 and Γ, is as

$$\bar{\tau} | \sigma^2, \Gamma \sim N(\bar{\tau}^*, D^*) \ .$$

This permits the computation of the predictive distribution of V given $X_1, \ldots, X_q$ and σ^2 and Γ which is

$$V|\sigma^2,\Gamma,X \sim N(W\bar{\tau}^*,\ W(\Gamma+D')W' + \sigma^2 I) \ . \tag{7.14}$$

Hence (7.14) can be used to generate an approximate predictive distribution with the insertion of estimates for σ^2 and Γ, as before.

For conditional prediction of $V^{(2)}$ given $V^{(1)}$ again standard normal theory is applied to (7.14) so that approximately

$$V^{(2)}|X,\ V^{(1)} \sim N(W^{(2)}\bar{\tau}^* + A_{21}A_{11}^{-1}(V^{(1)} - W^{(1)}\bar{\tau}^*),\ A_{22\cdot 1}) \tag{7.15}$$

where

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = W(\hat{\Gamma} + D^*)W' + \hat{\sigma}^2 I \ .$$

Hence

$$[V^{(2)}-W^{(2)}\bar{\tau}^*-A_{21}A_{11}^{-1}(V^{(1)}-W^{(1)}\bar{\tau}^*)]'A_{22\cdot 1}^{-1}[V^{(2)}-W^{(2)}\bar{\tau}^*-A_{21}A_{11}^{-1}(V^{(1)}-W^{(1)}\bar{\tau}^*)] \le \chi_\beta^2(p_2) \tag{7.16}$$

provides an approximate predictive region for $V^{(2)}$ given $V^{(1)}$.

One notes that this model and procedure makes many stringent assumptions and produces only approximate results--hence the model should be carefully checked and the results applied with caution. However, if this model is really appropriate it will produce the best results.

8. A Sample Reuse Approach for Conditional Prediction.

We now describe a data analytic method called Predictive Sample Reuse, Geisser (1974, 1975), which can be applied to conditional growth curve prediction, and does not require distributional assumptions. For convenience we make one small change in notation instead of V being the future vector observations we shall relabel

$$V = X_{q+1} = \begin{pmatrix} X_{q+1}^{(1)} \\ X_{q+1}^{(2)} \end{pmatrix} \ .$$

Suppose from the first q vectors, $X_1, \ldots, X_q$ each at the same p points we generate a predictor of $X_{q+1}^{(2)}$, say $\hat{X}_{(q)}^{(2)}$. Further suppose another predictor of $X_{q+1}^{(2)}$ is obtained, say $\hat{X}_{q+1}^{(2)}$, which depends only on the observed $X_{q+1}^{(1)}$. Finally we combine the two independently calculated predictors into a new predictor

$$\dot{X}_{q+1}^{(2)} = f(\hat{X}_{(q)}^{(2)}, \hat{X}_{q+1}^{(2)}; \Omega) \tag{8.1}$$

for $\Omega \in \mathfrak{J}$, $\mathfrak{J}$ being a specified class of matrices. An interesting case is

$$\dot{X}_{q+1}^{(2)} = \Omega\, \hat{X}_{(q)}^{(2)} + (I-\Omega)\, \hat{X}_{q+1}^{(2)} \tag{8.2}$$

where Ω is $p_2 \times p_2$ matrix such that Ω and $I-\Omega$ are both non-negative definite. Define

$$\dot{X}_{\alpha}^{(2)} = \Omega\, \hat{X}_{(q-1,\alpha)}^{(2)} + (I-\Omega)\hat{X}_{\alpha}^{(2)} \tag{8.3}$$

where $\alpha = 1, \ldots, q$ and $\hat{X}_{(q-1,\alpha)}^{(2)}$ is the predictor for $X_{\alpha}^{(2)}$ based on $X_1, \ldots, X_{\alpha-1}, X_{\alpha+1}, \ldots, X_q$ and of the same functional form as $\hat{X}_{(q)}^{(2)}$ and $\hat{X}_{\alpha}^{(2)}$ is the predictor of $X_{\alpha}^{(2)}$ based only on $X_{\alpha}^{(1)}$ and of the same functional form as $\hat{X}_{q+1}^{(2)}$. Further define a discrepancy measure

$$D(\Omega) = \sum_{\alpha=1}^{q} d(\dot{X}_{\alpha}^{(2)}, X_{\alpha}^{(2)}) \tag{8.4}$$

which is then minimized with respect to Ω within its given domain of definition. If $\hat{\Omega}$ is the unique solution then the final predictor is given as

$$\tilde{X}_{q+1}^{(2)} = \hat{\Omega}\, \hat{X}_{(q)}^{(2)} + (I-\hat{\Omega})\, \hat{X}_{q+1}^{(2)}\,. \tag{8.5}$$

If

$$d(\cdot,\cdot) = \sum_{\alpha=1}^{q} (\dot{X}_{\alpha}^{(2)} - X_{\alpha}^{(2)})'(\dot{X}_{\alpha}^{(2)} - X_{\alpha}^{(2)}) \tag{8.6}$$

and $m < p_1$ and $p_2 = 1$, i.e. Ω is 1×1, a solution for combining predictors, based on simple least squares predictors, appears in Geisser (1975) where $\hat{X}^{(2)}_{(q)} = W^{(2)}T_1$ and $\hat{X}^{(2)}_{q+1} = W^{(2)}(W^{(1)\prime}W^{(1)})^{-1}W^{(1)\prime}X^{(1)}_{q+1}$, though there only illustrated for $m = 2$. A general solution may easily be obtained for other forms of $\hat{X}^{(2)}_{(q)}$ and $\hat{X}^{(2)}_{q+1}$ when $m < p_1$ and p_2 is arbitrary as

$$\hat{\Omega} = [\sum_{\alpha=1}^{q}(X^{(2)}_{\alpha} - \hat{X}^{(2)}_{\alpha})(\hat{X}^{(2)}_{(q-1,\alpha)} - \hat{X}^{(2)}_{\alpha})'][\sum_{\alpha=1}^{q}(\hat{X}^{(2)}_{(q-1,\alpha)}-\hat{X}^{(2)}_{\alpha})(\hat{X}^{(2)}_{(q-1,\alpha)}-\hat{X}^{(2)}_{\alpha})']^{-1} \quad (8.7)$$

provided it exists and satisfies the constraints, The simplest way to achieve the solution (8.7) is to use the matrix differentiation technique described in De Waal and Nel (1978).

In summary we have described a low structure data crunching device which simulates the predictive process as best it can, given a complete lack of distributional assumptions. The method has its roots in cross-validation analysis.

9. Group Growth Curve Comparisons.

For standard distributional assumptions, we have already indicated in section 2 how certain traditional tests are executed for hypotheses on $C\tau D$. In particular for polynomial growth curves these tests not only require distributional assumptions but, also that each individual vector be measured at the same p points. Both of these assumptions can be relaxed if we admit randomization or permutation tests. Zerbe and Walker (1977) present a method of implementing such a permutation test to ascertain whether several group mean growth curves can be considered to be essentially one group or not, i.e. each group is identifiable by some label and one tests for the relevance of the group label. The test can also be specified for a particular subinterval of time.

Suppose N individuals comprise q groups so that the jth individual in the αth group is $X'_{\alpha j} = (X_{1\alpha j}, X_{2\alpha j}, \ldots, X_{p_{\alpha j}\alpha j})$. Now the data for each individual is fitted by least squares to a polynomial of degree $m_{\alpha j} - 1$ where $p_{\alpha j} > m_{\alpha j}$. Let m be the maximum of the $m_{\alpha j}$. Further $m_{\alpha j}$ is also a polynomial of degree m by virtue of augmenting the residual $m - m_{\alpha j}$ terms of the polynomial with zero coefficients.

Let the polynomial fitted to $X'_{\alpha j}$ be represented as $x_{\alpha j}(t)$ and be considered to have a population average of $\tau_\alpha(t)$. Further assume that we wish to test whether the hypothesis that $\tau_\alpha(t) = \tau(t)$ $\alpha = 1, \ldots, q$ for all t over some interesting interval of time $t_1 \leq t \leq t_2$. Hence the authors propose the following analysis: Let

$$x_{\alpha\cdot}(t) = N_\alpha^{-1} \sum_j x_{\alpha j}(t) , \quad x_{\cdot\cdot}(t) = N^{-1} \sum_\alpha \sum_j x_{\alpha j}(t) \tag{9.1}$$

and

$$\int_{t_1}^{t_2} (x_1(t) - x_2(t))^2 dt \tag{9.2}$$

be defined as a measure of the squared distance between two curves and arrange the data in the following table:

ANOV TABLE II

Source	SS	DF
Groups	$B = \sum_\alpha N_\alpha \int_{t_1}^{t_2} (x_{\alpha\cdot}(t) - x_{\cdot\cdot}(t))^2 dt$	q-1
Within Groups	$E = \sum_\alpha \sum_j \int_{t_1}^{t_2} (x_{\alpha j}(t) - x_{\alpha\cdot}(t))^2 dt$	N-q

It would appear that an F ratio statistic, say,

$$F_0 = [(N-q)B]/[E(q-1)] \tag{9.3}$$

should be sensitive to alternatives in relation to the degree that

$$\sum_{\alpha} N_{\alpha} \int_{t_1}^{t_2} (\tau_{\alpha}(t) - \tau(t))^2 dt \tag{9.4}$$

departs from zero. The F ratio is computed for every permutation of the N individual curves such that there are N_{α} in the αth group. The significance level β is the number of such F ratios that are at least as large as the observed permutation given by F_0. Since the number of such ratios may be prohibitive to compute it is suggested that β be estimated by choosing at random d of these permutations and noting the number f which are at least as large as F_0 and using the binomial distribution to place confidence limits about β.

10. Concluding Remarks.

Growth curve analysis as presented by most statisticians has, until rather recently, generally stressed testing and estimation of the set of parameters τ. We have attempted to shift the focus so as to emphasize more strongly, prediction, and yet discuss the traditional concerns. There are several reasons for the predictivistic point of view. The first is that an investigator is often more concerned with prediction than testing and estimation even if the growth curve model could, by some stretch of the imagination, be assumed a true representation of the physical process underlying the responses. Secondly, it is quite clear that growth curve models do not in this sense provide such an exact physical specification. They are basically statistical paradigms that are particularly convenient and useful for vastly complex phenomena about which knowledge is often incomplete, fuzzy and generally lacking in the fundamental relationships. Lastly, prediction involves the entity that investigators actually measure--the response itself. Thus predictions can be, to a degree, validated by further investigation which is not the case with those hypothetical constructs--the parameters of the model, unobservable as they are.

Hence testing, e.g. two groups are the same or differ in regard to their growth curve parameters, is really a selection problem in that we should choose the alternative that enhances prediction from one or both of the groups. The conventional .05 level for rejecting a null hypothesis is probably not very good for this purpose. That the estimation of τ can be of some interest e.g. estimating the differential growth rate of two groups whose growth curves are both approximately linear, we do not deny. But again this is, in a sense, associated with, and finally subordinate to prediction because the differential growth rate is really useful for describing where a future response from a random unit from one group will be, compared to one from another group. This of course, can be ascertained without recourse to estimating τ.

This work was supported in part by a grant from the National Institute of General Medical Sciences.

References

Bowden, D. C. and Steinhorst, R. K. (1973). Tolerance bands for growth curves. Biometrics, 29, 2, 361-371.

Box, G. E. P. (1950). Problem in the analysis of growth and wear curves. Biometrics, 6, 362-389.

Cole, J. W. L. and Grizzle, J. E. (1966). Applications of multivariate analysis of variance to repeated measurements experiments. Biometrics, 22, 4, 810-828.

Collier, R. O., Baker, F. B., Mandeville, G. K., and Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika 32, 339-353.

Danford, M. B., Hughes, H. M., and McNee, R. C. (1960). On the analysis of repeated-measurements experiments. Biometrics, 16, 547-565.

De Waal, D. J. and Nel, D. G. (1978). Parametric Multivariate Analysis, University of the Orange Free State. Bloemfontein.

Fearn, T. (1975). A Bayesian approach to growth curves. Biometrika, 62, 1, 89-100.

Geisser, S. (1964). Posterior odds for multivariate normal classification. J. Roy. Statist. Soc. Ser. B, 26, 69-76.

Geisser, S. (1965a). A Bayes approach for combining correlated estimates, Journal of the American Statistical Association, 60, 602-607.

Geisser, S. (1965b). Bayesian estimation in multivariate analysis. Ann. Math. Stat., 36, 150-159.

Geisser, S. (1966). Predictive discrimination. Multivariate Analysis, P. Krishnaiah, editor, Academic Press, New York, 149-163.

Geisser, S. (1970). Bayesian analysis of growth curves, Sankhyā A, 32-53-64.

Geisser, S. (1971). The inferential use of predictive distributions. Foundations of Statistical Inference. J. Godambe and D. Sprott, eds. Rinehart and Winston, 456-469.

Geisser, S. (1974). A predictive approach to the random effect model. Biometrika, 61, 101-107.

Geisser, S. (1975). The predictive sample reuse method with applications. J. Amer. Statist. Assoc., 70, 350, pp. 320-328.

Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. J. Roy. Stat. Soc., Series B, 25, 368-376.

Geisser, S., and Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. Ann. Math. Statist., 29, 885-891.

Geisser, S. and Kappenman, R. F. (1971). A posterior region for parallel profile differentials. Psychometrika, 36, 1, 71-78.

Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.

Grizzle, J. E. and Allen, D. M. (1969). Analysis of growth and dose response curves. Biometrics, 357-381.

Halperin, M. (1961). Almost linearly-optimum combination of unbiased estimates, Journal of the American Statistical Association, 56, 36-43.

Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. Ann. Inst. Stat. Math., 18, 75-86,

Lee, J. C. (1977). Bayesian classification of data from growth curves. S. African Stat. J. 11, 155-166.

Lee, J. C. and Geisser, S. (1972). Growth curve prediction. Sankhyā, A, 34, 393-412.

Lee, J. C. and Geisser, S. (1975). Applications of growth curve prediction. Sankhyā, A, 37, 239-256.

Lee, Y. S. (1972). Some results on the distribution of Wilks Likelihood-ratio criterion. Biometrika, 95, 3, 649-664.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). J. R. Statist. Soc. B, 34, 1-41.

Nagel, P. and DeWaal, D. (1978). Bayesian classification and estimation of growth curves. S. African Stat. J. (in press).

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326.

Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. Biometrika, 46, 49-58.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika, 52, 447-458.

Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis. Multivariate Analysis, P. Krishnaiah, editor. Academic Press, 87-103.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its applications to measurement of signals. Proc. 5th Berkeley Symp. 1, 355-372.

Rao, C. R. (1968). A note on a previous lemma in the theory of least squares and some further results. Sankhyā, A, 30, 245-252.

Rao, C. R. (1969). A decomposition theorem for vector variables with a linear structure. Ann. Math. Statist., 40, 1845-1849.

Rao, C. R. (1972). Recent trends of research work in multivariate analysis. Biometrics, 28, 1, 3-22.

Rao, C. R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. Biometrics, 31, 2, 545-554.

Schatzoff, M. (1966). Exact distribution of Wilks' likelihood ratio criterion. Biometrika, 53, 347-358.

Smith, A. F. M. (1973). A general Bayesian model. J. R. Stat. Soc. B, 35, 67-75.

Wilson, K. (1975). The sampling distribution of conventional, conservative, and corrected F-ratios in repeated measurements designs with heterogeneity of covariance. Journal of Statistical Computation and Simulation, 3, 201-215.

Zerbe, G. O. and Walker, S. H. (1977). A randomization test for comparison of groups of growth curves with different polynomial design matrices. Biometrics, 33, 4, 653-659.